Indian J. Phys. **64B** (5), 377-385 (1990)

A fresh look on DNA double helix : Part I. A critical analysis of two base pair schemes. A rule and role of Watson-Crick ⇔ Hoogsteen base pair in DNA polymorphic transitions*

Bankim Das, Subhasis Chakravarty and Asok Banerjee+
Biophysics Department, Bose Institute, Calcutta-700 054, India

Received 7 August 1989, accepted 22 May 1990

Abstract : Two important base pair schemes are critically re-examined to account for dynamic aspects observed in transition topology in DNA double helix. Sterically viable conformation of the base pair scheme in parallel and antiparallel DNA double helix emerges a generalised selection rule (Table 1) about the most probable conformation in DNA. The rule predicts that the simple allowance in plectonomically wound double helix whether made of parallel or antiparallel chains to undergo syn ⇔ anti to base pairs takes out major restrictions for a well base paired double helix (either Watson-Crick or Hoogsteen). The rule also highlights a rational explaination as to why the probability of occurence of antiparallel compared to parallel chain double helices is much higher in tune with the observed data. Flexibility inherent in the base pair schemes, necessarily in vivo, needs syn ⇔ anti or vice versa transition in conformation for function and base pair preservation (Figure 1). Classical W-C model has no conflict with this model and is a member of this generalised model and it explains nicely the observed data for DNA conservatism, different diameter at different stretches, melting point, induced fit, overlapping gene phenomena more effectively in DNA paradigm.

It is interesting that Gueron *et al* (1989) have concluded that Hoogsteen pairing is the stable form of GC pairs in B-DNA at low pH, and Hoogsteen structure is formed transiently at neutral pH. Recently, Vande Sande *et al* (1988) and Pattaviraman (1986) have also found the existence of parallel double helix in DNA paradigm in tune with our predicted rule.

Keywords : DNA transition topology, DNA polymorphism, DNA base pairs, DNA parallel and antiparallel helix, selection rule in DNA double helix.

PACS No : 87.15.BY

*Essentials of this paper were presented as invitee by NSF, USA in the International Symposium on "6th. Conversation Biomolecular Stereodynamics", Suny Albany, June 06-10, 1989.

+To whom correspondance is to be made.

1. Introduction

Structural polymorphism of DNA have been studied for nearly thirty years by fibre diffraction techniques and by X-ray analysis of single crystals of deoxyoligonucleotides in the past few years. It has been observed that DNA manifests different polymorphic forms and these forms and their inter transitions depend primarily on (a) internal factors such as sequences, compositions etc., (b) external factors such as relative humidity and salt concentration, temperature, pH etc. and (c) topological stress factors (Chakravarty *et al* 1988).

These different important polymorphic forms important in living systems are related with the structural basis of drug-nucleic acid/protein-nucleic acid interaction in spatial reference to emerged new structural features/models that may be involved in vivo interactions and consequently in function. DNA binding proteins recognize specific base sequences in DNA by tight binding to these sites. These non-covalent associations play a key role in regulating gene-expression—a process involved in information transfer, from DNA to RNA and then RNA to proteins. The first step in this readout process involves tight binding of protein (RNA polymerase enzyme) to specific DNA sites called 'promoters'. The details of these interactions are complex. Promoters can be conformationally altered, premelted DNA regions, arise as a consequence of structural phase transition in the polymer (Sobell *et al* 1983b). Different regions of DNA undergo transition at different specific time under torsional stress factors. These polymorphs A-, B-, C-, D-DNA in the right-handed and Z-DNA in the left-handed family are likely to be involved in vivo, and therefore, a rational attempt to an understanding of the specific nature of binding of proteins associated with the stabilization of these right and left-handed helices are very much needed. The interaction between Z-DNA and specific binding proteins have been studied using anti Z-DNA anti bodies as a model system (Nordheim *et al* 1981). The details of these interactions are complex, mostly unknown and a major unclear area in molecular biology. But how is such a precise interaction achieved ? Does the recognition require the Watson-Crick B helix, or other helices (right or left-handed) with non Watson-Crick base pair or could there be some other DNA structures, (promoters) with more versatile *base pair scheme* that signals these interactions ?

In this part (part I) of the paper on *a fresh look on DNA double helix*, we like to revisit the two important existing static base pair schemes and focus on dynamic aspects associated with W-C ⇔ Hoogsteen base pair in DNA double helix transition and its *inseparable, indispensible interplay* in polymorphic forms and transition topology in DNA paradigm.

2. Materials and methods

The Watson-Crick base pair, Hoogsteen base pair, evidence of Hoogsteen and W-C paired double helix and flip-flop between W-C and Hoogsteen base pair :

There is a wealth of crystallographic data on base pair schemes (Sobell *et al* 1983b). Figure 1 shows two most important prevalent base pair schemes as observed (The Watson-Crick and Hoogsteen) from crystal structure analysis. The

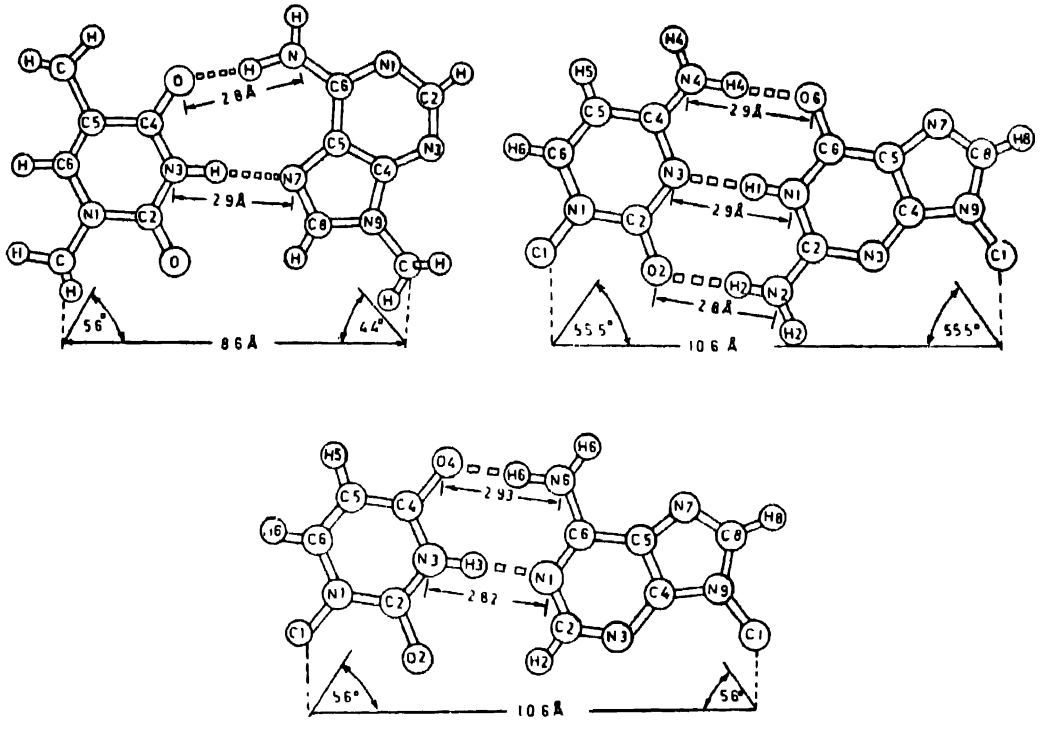

Figure 1. The base pairing scheme : Hoogsteen and Watson-Crick base pairing for A-T and G-C with bondlength, bondangle and hydrogen bonded distances.

importance of the W-C complementary base pairing scheme should not overshadow alternative types of base pairs in DNA, one of which is Hoogsteen pairing, the presence of which has been established in crystals of oligonucleotides bound to antibiotics and its possible existance in solution. Variability in DNA conformation play an important biological role and thus W-C ⇔ Hoogsteen base pair poses an interesting factor in inducing such changes. Let us revisit these base pairs with a fresh look. The W-C base pair is characterised by a pseudo dyad axis, hydrogen bonding between purine and pyrimidines via N1...N3, N6...O4 and Hoogsteen pair by hydrogen bonding of N7...N3, N6...O4. The other interesting difference between these two types of pairing is that the average distance between glycosidic points in case of W-C is 10.6 Å while it is 8.6 Å in case of Hoogsteen pairing. The glycosidic bonds make angles with the line joining them in case of W-C is symmetrical and average is 55.5° on both sides while in case of Hoogsteen it is asymmetric and the corresponding angles are 56° and 44° respectively (Figure 1). The interconversion of W-C and Hoogsteen base pair will therefore need a rotation

of purine base of 180° about glycosidic bond and a shearing motion of bases before a pairing is stabilized ; and a double helix of W-C base pair will be larger in diameter compared to one with Hoogsteen base pair (Figure 2).

There are generally two grooves in a DNA double helix. Inside these grooves are edges of the base pairs consisting of donors as well as acceptors for hydrogen

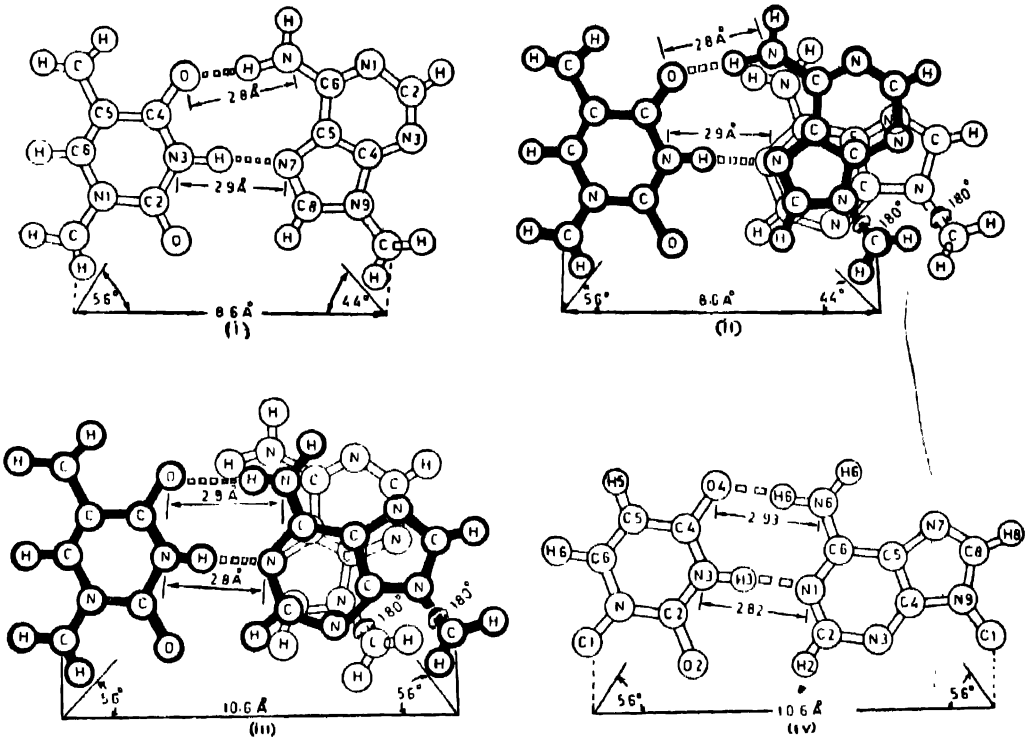

Figure 2. The schematic representation of transition of Hoogsteen ⇔ W-C base pairing. The flipping mechanism about N_9 - - - C_1 bond is shown.

bonding interactions with various ligands including drugs, carcinogens and mutagens. The base pair scheme will have a distinct impact in these groove pattern and the teething disposition of the edges of the base pairs will affect any interactions with it in vivo (Wang *et al* 1984). Recently structural features observed in the complex of two quinoxaline antibiotics with DNA d(CGTACG) hexamer show GC and AT base pairs flanking the quinoxaline rings of the drug molecule in Hoogsteen geometry (Figure 2) while the central base pairs are in the conventional W-C type. This result points out that though G-C Hoogsteen base pairs are less stable than W-C base pair, the base pair derives its stability via Van der Waals (including stacking) interaction between the quinoxaline rings and the sugar phosphate backbones associated with the Hoogsteen base pairs.

In this complex, four of the purines are in the syn conformation and four in the *anti* conformation. It is known that purines adopt syn conformation readily and this is one of the basic structural elements underlying the formation of the left handed Z-DNA structure in which every alternating residue is in syn conformation. In fact, alternative DNA conformations and models have been proposed with purines in the syn conformation and Hoogsteen base pairing (Drew and Dickerson 1982, Pulley *et al* 1985 and Banerjee *et al* 1984 and Lalwani 1987). A right handed DNA double helix with alternating (A-T) with syn adenine and anti thymine paired together with Hoogsteen geometry can be modelled Wang *et al* 1987.

This right handed Hoogsteen base paired double helix is overwound relative to B-DNA with a slimmer helix diameter. This model has features consistent

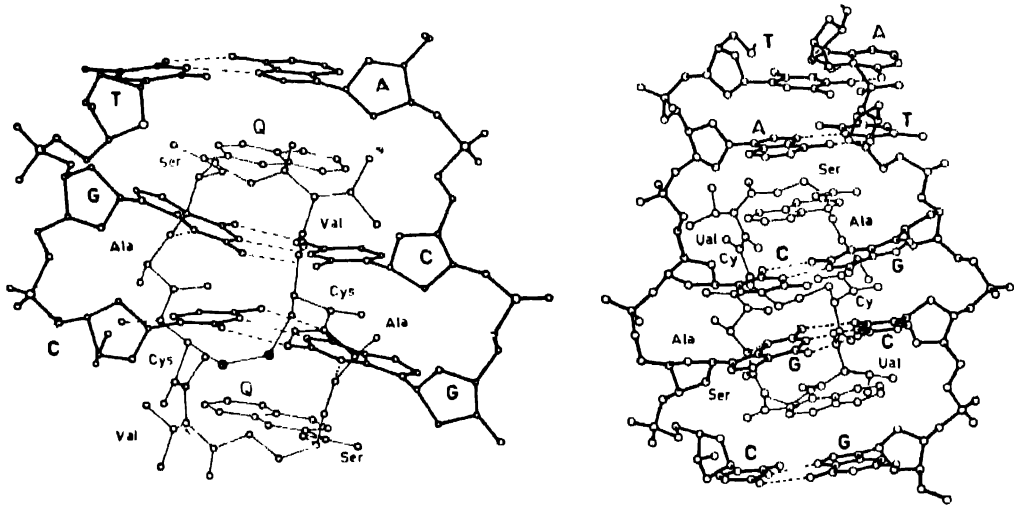

Figure 3. Crystallographic evidence of co-existance of both W-C and Hoogsteen base pairing in a mini DNA double helix complexed with Triostien-A drug. (a) Major groove view and (b) Minor groove view.

with those of D-DNA (Arnott *et al* 1974, Mahendrasingam *et al* 1983 and Chandrashekharan *et al* 1984) but no definite correlation has yet been established.

Observed diffraction patterns of DNA strongly suggest other stable alternative DNA conformations. This is due to the fact that DNA is conformationally active (Sobell *et al* 1983b) via non-linear inversion of sugar moiety (Banerjee and Sobell 1983 and shearing motion of base pairs and flip-flop mechanism of the base pairing W-C ⇔ Hoogsteen in DNA double helix via DNA breathing (Wang *et al* 1987). This interconversion of W-C to Hoogsteen and vice versa is achieved via flipping of base mostly purine about its glycosidic bond within an intact double helix plectonomically wound (most likely in the extended intercalative conformation of DNA).

It is apparent that there is restriction or rule on feasibility of base pairs, governing a sterically viable DNA double helix and the effect of the rule and role of flip-flop (off-on) of W-C ⇔ Hoogsteen base pair scheme on DNA structure, topology and transition is inseperable. It is highly unlikely that facility is not exploited by DNA in biological system during its interaction in vivo.

3. Results and discussion

(a) DNA base pair flexibility and transition topology :

A critical model analysis of double helical DNA structures reveal (*all view from Minor Groove*)-the possible double helices abiding selection rule (Table 1).

Table 1. The selection rule governing the base paired DNA double helix.

Nature of base pair	Nature of the chain (3′-5′) in the helix					
	↓ ↑			↑ ↑		
	Anti	Syn	Syn-anti	Anti	Syn	Syn-anti
Watson-Crick	+	+	+	-	-	+
Hoogsteen	-	-	+	-	-	+

\+ → possible
\- → not possible
↑ ↓ → anti parallel chain arrangement.
↑ ↑ → parallel chain arrangement.

The emerged fact is that 'syn-anti' or 'anti-syn' conformations are possible in both parallel and anti-parallel chains having Watson-Crick and Hoogsteen base pairs.

Therefore, it may be argued that syn-anti alteration in DNA double helix (at different stretches) is a pre-requisite for flexibility of base pairing into Watson-Crick or Hoogsteen and vice-versa, in case of functional necessity. The probability of either is exactly 50% except steric hindrence, if any. Thus possibility of existance of both kinds of base pairing in a double-helix not only increases its versatility and specificity as regards binding with ligands, but these base pairings modulate groove pattern enormously for specific interaction with proteins and drugs in vivo. If chains are antiparallel, a stereochemically feasible Watson-Crick base pair is also possible with anti conformation on both chains with chain interchange relative to conventional W-C double helix. This constitutes a non conventional DNA helix with interesting groove features (Hopkins 1981 and Sobell *et al* 1983a).

It seems very interesting that the mode for flexibility of 'anti' ⇔ 'syn' conformation about glycosidic bond (χ) wtihin a double helix is inherent (via

likely intercalation type geometry) and assists for effective interconversion of W-C or Hoogsteen base pairing in the conformational domain of DNA double helix ; and DNA chains whether parallel or anti-parallel seem to be of lesser importance in this regard ; nevertheless, the preference for anti parallel chains in double helix is in fact observed in 'nearest neighbour frequency' data (Josse *et al* 1961) and is nicely predicted by our selection rule (Table 1).

The impact of the rule on genetic code is worth to mention here. Existance of Watson-Crick and Hoogsteen base pairings are established in biology and their importance can not be neglected in view of their likely influence in codon-anticodon interaction in genetic coding at Wobble pairing. The answer to the question of degeneracy in genetic code might have a deep root in flip-flop/on-off mechanism of base pair via syn/anti alteration in interaction pathways where sequence remains the same but conformation is the determining factor. Model studies show that the DNA double helix of any of the established polymorphic forms are capable of flip or flop (syn/anti) within a stretch of double helix (DNA or RNA or DNA RNA hybrid) without complete disruption of hydrogen bonded plectonomically wound double helical strands at all.

This mechanism might be involved in restricting the degeneracy by specific interaction of the W-C or Hoogsteen type of base in the 3rd position of the Wobble pair. We predict that in most DNA polymorphism (in different plectonomical forms, A-, B-, C-, D-, E-, Z- and other forms to reveal in future) the basic mechanism lies in this *syn-anti* conversion of bases to these two basic base pairs W-C and Hoogsteen type and DNA double helix must follow the selection rules laid down here for base pair preservation, a basic necessary criteria in a double helix.

It is worth to mention that if W-C base pairing is prevented (e.g. with poly A substituted in the two position of adenine by methyl groups), poly A.poly U may occur in Hoogsteen form (Saenger 1984, Hakoshima *et al* 1981). H-form of DNA with homopurine-homopyrimidine repeats under superhelical stress or acid pH appears to be a triple helix structure with a W-C double helix associated with a homopyrimidine loop by Hoogsteen base pairs (Lyamichev *et al* 1987, Mirkin *et al* 1987).

It is interesting that Gueron *et al* (1989) have concluded that Hoogsteen pairing is the stable form of GC pairs in B-DNA at low pH, and Hoogsteen structure is formed transiently at neutral pH. Recently, Vande Sande *et al* (1988) and Pattaviraman (1986) have also found the existance of parallel double helix in DNA paradigm in tune with our predicted rule.

(b) A few interesting features that our model explains :

1. Different diameter of DNA at different stretches along DNA length for protein, ligand binding effectively with non monotonus groove pattern and

major surface topology change due to conformation in either W-C or Hoogsteen geometry. The overall shape change will be more pronounced than sequence change.

2. Confirms conservatism of DNA double helix ; this model helix only disrupts to seperate strands in case of necessity. Otherwise performs function via groove modulation by syn ⇔ anti conversion.
3. Nearest neighbour base frequency data, hitherto unexplained properly is nicely explained. The anti-parallel double helix is preferred over parallel double helix simply by base pair rule (Table 1).
4. Complete conformity with classical W-C double helix ; classical DNA double helix is a particular case of this model.
5. Different binding affinity for enzymes at definite site depending on the teething disposition of DNA due to base pair conformation either in syn or anti. Besides sequence, this keeps open an extra privilege for enzyme or substrate for specific binding. This model has privilege in overlapping gene phenomena in genetic expression.
6. Probable alternative physical explanation for non sharp melting point of the DNA where plectonomical double helix does not completely disrupt into strands rather base pairs disrupt (via syn anti) to W-C or Hoogsteen.
7. This double helix DNA model changes topology of DNA receptor for different substrates, enzymes, ligends etc. to fit in (induced fit) via flip-flop of bases.

4. Conclusion

(1) Simple allowance to undergo SYN ⇔ ANTI to the base pairs takes out major restriction (Table 1) for an intact preservation of a well base paired (WC, HS) plectonomically wound double helix whether parallel or anti parallel chains.

(2) A rational explanation to the unanswered questions as to why parallel chain double helix is less prevalent compared to antiparallel in vivo as per data (Aurther Kornberg) ; our Table 1, shows probability is higher for antiparallel chains though it does not exclude possibility of parallel double helix.

(3) Syn-Anti *alternation* in a DNA double helix is not essential but allowence of syn ⇔ anti *alteration* (conversion) is the necessary and sufficient condition for basic kinds of conformational changes and transition topology in DNA double helix.

(4) In a plectonomical double helix syn ⇔ anti conversion is the prerequisite for base pair preservation both for W-C or Hoogsteen base pairing. This means that classical W-C double helix (anti parallel chains, both sugars in anti conformation, B-DNA) has no conflict with this model and is a particular case of our generalised model.

(5) The importance of W-C complementary base pairing scheme along with existing alternative types of base pairings are better understood through this exercise.

Acknowledgment

Our thanks are due to CSIR/DST, Govt of India for financial help. Thanks are also due to all our colleagues, students who supported, argued, and gave us inspiration.

References

Arnott S, Chandrasekharan R Hukins D W L Smith P J C and Walts L 1974 *Mol. Biol.* **88** 523

Banerjee A 1984 *Proc. Int. Cell. Biology* (Japan) **5250** 566

Banerjee A and Sobell H M 1983 *J. of Biomolecular Structure and Dynamics* 1 253

Chakravarty S, Lalwani R and Banerjee A 1988 *Indian J. Phys.* **62B** 137

Chandrasekharan R, Arnott S, Banerjee A, Smith S C, Leslie A G W and Puigjaner 1980 *Fibre diffraction methods ACS Symposium* **Series 141** (American Chemical Society) 483

Drew H R and Dickerson R E 1982 *The EMBO J.* **1** 663

Gueron M, Charretier E, Kochoyan M and Leproy J L 1989 *Book of Abstracts : 6th. Conversation in Biomolecular Stereodynamics* p 122

Hakoshima T, Fukui T, Ikehara M and Tomita K I 1981 *Proc. Natl. Acad. Sci.* **78** 7309

Hopkins R C 1981 *Science* **211** 289

Josse J, Kaiser A D and Kornberg A 1961 *Biol. Chem.* **236** 864

Lalwani R Chakravarty S and Banerjee A, 1987 *Acta. Cryst.* A43 (Suppl) 02. 5 C3

Lyamichev V I, Mirkin S M and Frank-Kamenetskii M D 1987 *J. Biomolecular Structure and Dynamics* **5** 275

Mahendrasingam A, Rhodes N J, Goodwin D C, Nave C, Pigram W J, Fuller W, Brahnis J and Veryne J 1983 *Nature* **301** 535

Mirkin S M, Lyamichev V I, Drushlyak K N, Dobrynin V N, Filippor S A and Frank-Kamenetskii M D 1987 *Nature* **330** 495

Nordheim A, Pardue M L, Lafer E M, Moller A, Stollar B D and Rich A 1981 *Nature* **294** 417

Pattaviraman N 1986 *Biopolymers* **25** 1603

Pulley blank D E, Haniford D B and Morgen A 1985 *CELL* **42** 271

Saenger W 1984 *Principles of Nucleic Acid Structure* (Berlin : Springer-Verlag)

Sobell H M, Banerjee A, Edward D, Lozansky, Guo Pingzhou G and Chen C 1983a *Structure and Dynamics : Nucleic Acids and Proteins* eds Enrico Clementi and R H Sharma (New York : Adenine)

Sobell H M, Sakore T D, Jain S C, Banerjee A, Bhandary K K, Reddy B S and Lozansky E D 1983b *Cold Spring Harbar Symposia On Quantitative Biology* Vol-XLVII

Vande Sande J H, Ramsing N B, German M W, Horest W E L, Kalisch B W, Kitzing E V, Don R T, Clegg R C and Jovin T M 1988 *Science* **241** 551

Wang A H J, Ughetto G, Quigley G H, Hakoshimen T, Vander Marel G A, Van Boom J H and Rich A 1984 *Science* **225** 1115

Wang A H J, Ughetto G, Quigley G J and Rick A 1987 *Biochemistry* **26** 1152